

Texmix: An automatically generated news navigation portal

Morgan Brehinier¹ and Guillaume Gravier²

¹ INRIA Rennes,
`morgan.brehinier@inria.fr` ,
² CNRS-IRISA,
`guillaume.gravier@irisa.fr`

Abstract. The Texmix demonstration presents an original interface to navigate a collection of broadcast news shows, exploiting speech transcription, natural language processing and image retrieval techniques. Navigation is performed through keywords search or through time or through maps, with links automatically created either between reports to follow the story or to the Web to know more about a story, a person or a fact. Image search technology is also integrated to find portions of the collection with similar images. We also present two original features to dynamically access videos, namely dynamic summary and geotagging ...

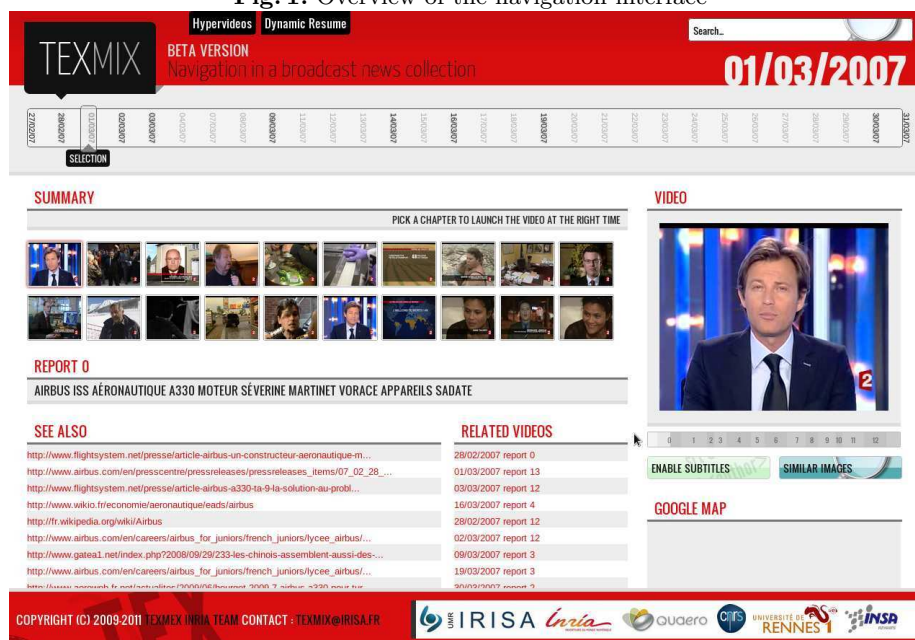
Keywords: HTML5, hypervideo, topic segmentation, speech to text, named entities, similar images, dynamic resume, geotagging

1 Introduction

The gradual migration of television from broadcast diffusion to Internet diffusion offers tremendous possibilities for the generation of rich navigable contents. However, it also raises numerous scientific issues regarding delinearization of TV streams and content enrichment. In this demonstration, we illustrate how speech in TV news shows can be exploited for delinearization of the TV stream. In this context, delinearization consists in automatically converting a collection of video files extracted from the TV stream into a navigable portal on the Internet where users can directly access specific stories or follow their evolution in an intuitive manner. The TEXMIX demonstration illustrates the result of the delinearization process and enables user to experience various navigation modes within a collection of news shows. An illustration of the interface is provided Figure 1. The demonstration can be accessed online at <http://texmix.irisa.fr>.

Structuring a collection of news shows requires some level of semantic understanding of the content in order to segment shows into their successive stories and to create links between stories in the collection, or between stories and related resources on the Web. Spoken material embedded in videos, accessible by means of automatic speech recognition, is a key feature to semantic description of video contents. At IRISA/INRIA Rennes, we have developed multimedia

Fig. 1. Overview of the navigation interface



content analysis technology combining automatic speech recognition, natural language processing and information retrieval to automatically create a fully navigable news portal from a collection of video files [1].

The purpose of this demonstration, aside from exploiting technologies developed in our research group, is to provide a new experience from watching videos on the Internet. Exploiting HTML5 features, dynamic content is provided as the video plays, and allow the watcher to have additional information and to interact during his watching, being no longer passive. We also provide new search services based on the technologies developed.

In this summary, we describe the various features of the demonstration, providing brief insight on the underlying technology.

2 Timeline navigation

An important feature to access and navigate a collection of news bulletin is time. A timeline is used to locate in time all information that you can access. It is also used to locate in time information that are displayed by positioning the cursor accordingly.

3 Link creation

News shows are made of successive items which are usually introduced by the anchor speaker and developed in some story. The first step to structuring a collection therefore consists in segmenting each show into its constituting items. This segmentation step is done using statistical topic segmentation method which detect topic changes in the transcription, inspecting the distribution of words. Topic segmentation enables to automatically generate a table of contents of each show. Furthermore, keywords are automatically extracted from the transcript to characterize each story. Exploiting the keywords and the transcripts, links to the Web and to related stories in the collection are created and displayed, thus offering navigation capabilities to follow the evolution of an event or to know more about it.

4 Named entities

Named entities, that is proper names and location names, are also automatically detected using machine learning algorithms and used to locate news items and to dynamically provide additional information while playing the video. For instance, a Wikipedia pop-up is displayed whenever a person's name is uttered, with a link to the corresponding wikipedia page for more information. Similarly, locations are dynamically displayed on a Google Map as they occur, thus facilitating the localization of news stories.

5 Similar images

Spoken language provides crucial information but so do images! Indeed, the same image is often used to illustrate a story throughout time. Based on our image retrieval technology [2], we included a 'similar image' search feature in the news portal, as illustrated Figure 2. In this second illustration, we search for images similar to a graphic depicting the voting intentions for the 2007 French presidential election. This results in similar graphics at different dates, making it possible to follow the evolution of the poll in time.

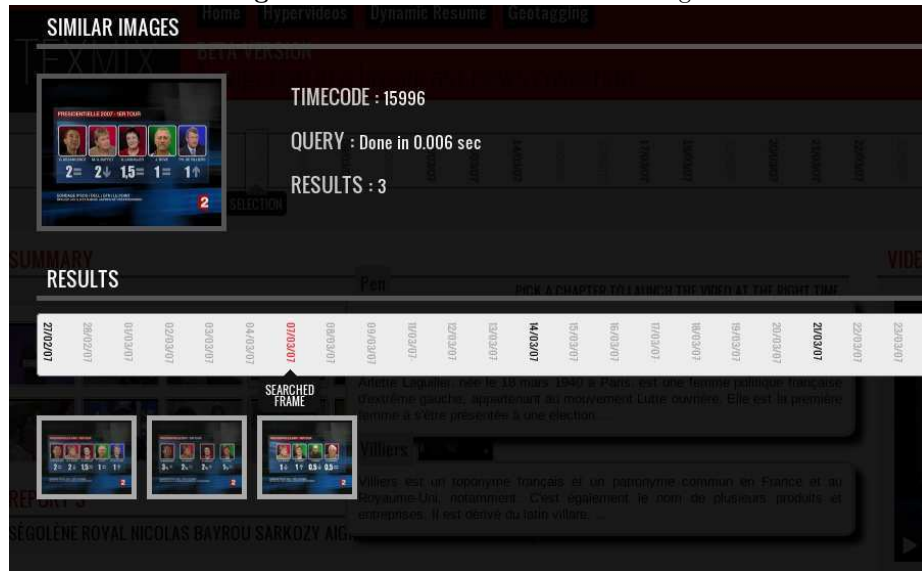
6 User services

Based on the previous technologies, we developed two services for the user.

6.1 Dynamic summary

Based on the keywords extraction technology, the dynamic summary offers the possibility to catch-up news on a specific topic in a very quick way. The user starts by selecting the period of time he is interested in, and then selects a keyword related to this period. Those two actions result in a dynamic resume of the topic, by sequencing the thirty first seconds of each story related to the keyword.

Fig. 2. Results of a search for similar images



6.2 Geotagging

In the same way 'dynamic resume' offered a navigation based on time and keywords, geotagging is a way to look for news stories related to their location. For this feature, the user starts by selecting the time period he is interested in, and markers are then displayed on a Google Map, showing where the news stories of this time period occurred. The user can then zoom in or click on a marker in order to view the news bulletin related to the place he chose.

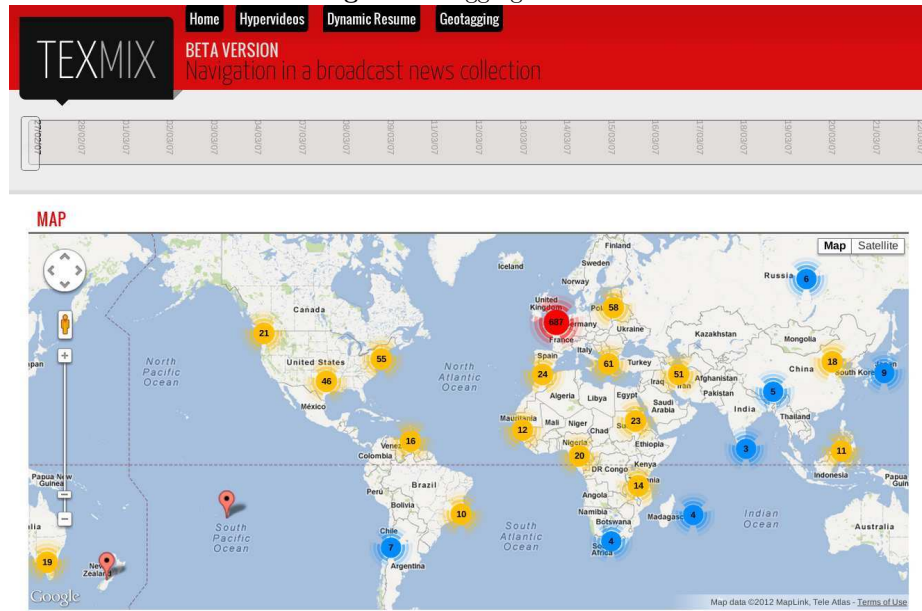
7 Technology

This interface has been developed in HTML5, CSS and Javascript. Those technologies allowed us to develop this multimedia portal without relying on proprietary software solutions. The synchronization of the contents and the videos is based on the popcorn.js library (www.popcornjs.org). Therefore, this demonstration is accessible on the Internet with any recent web-browser supporting HTML5 technologies.

8 Conclusion

This unique combination of speech and language technology and multimedia retrieval offers countless opportunities to automatically generate Web content from TV streams, opening the door to a fully connected multimedia Web.

Fig. 3. Geotagging interface



9 Acknowledgements

This work was partially funded by the Quaero project and by EIT ICT Labs via the OpenSEM project.

References

1. Guillaume Gravier, Camille Guinaudeau, Gwenole Lecorve and Pascale Sebillot. Exploiting speech for automatic TV delinearization: From streams to cross-media semantic navigation. In *Journal of Image and Video Processing*, 2011, 2011:689780.
2. Herve Jegou, Mathias Douze and Cordelia Schmidt. Product quantization for nearest neighbor search. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 33(1):117128, 2011.